# BDA hw2

### 2020311198 Dongook Son

## Contents

## [1] Chapter 2 Exercise Questions

**1.**

The prior distribution is as follows,

$$\pi(\theta) \propto \theta^{4-1}(1-\theta)^{4-1}$$
$$= \theta^3(1-\theta)^3$$

The likelihood function follows,

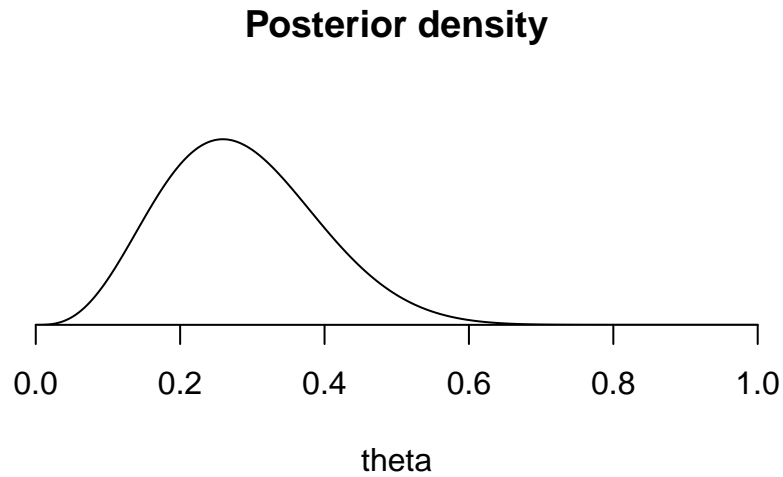$$P(obs|\theta) = \sum_{k=0}^{2}\binom{10}{k}\theta^k(1-\theta)^{10-k}$$

The posterior density for $\theta$ follows,

$$P(\theta|\pi, obs) = \frac{\pi(\theta)P(obs|\theta)}{\int \pi(\theta)P(obs|\theta)d\theta}$$
$$\propto \pi(\theta)P(obs|\theta) \cdots \text{denominator: constant}$$
$$= \theta^3(1-\theta)^{13} + 10\theta^4(1-\theta)^{14} + 45\theta^5(1-\theta)^{12}$$

```
theta <- seq(0,1, 0.001)
range <-theta^3*(1-theta)^{13} +
        10*theta^4*(1-theta)^{14} +
        45*theta^5*(1-theta)^{12}
plot(theta, range, type = 'l',
     main = 'Posterior density',
     ylim = c(0, max(range)*1.2),
     ylab = '', yaxt='n', bty='n', xaxs='i', yaxs='i')
```



**Posterior density**

**2.**

We can set up the variable in the following way

T: case of getting tails
TT: case of getting two tails
H: case of getting heads
N: number of spins until H
$\pi$: probability of H

We can derive the pmf of $N|\theta$.

| $N|\theta$ | mass |
|---|---|
| 1 | $\pi$ |
| 2 | $\pi(1-\pi)$ |
| 3 | $\pi(1-\pi)^2$ |
| 4 | $\pi(1-\pi)^3$ |
| 5 | $\pi(1-\pi)^4$ |
| $\ldots$ | $\ldots$ |
| k | $\pi(1-\pi)^{k-1}$ |

$E(N|\theta)$ can be derived easily by calculating the following equation.

$$E(N|\theta) - (1-\pi)E(N|\theta) = \pi + (1-\pi)\pi + \ldots$$

$$E(N|\theta) = 1 + (1-\pi) + (1-\pi)^2 + \ldots = \frac{1}{1-(1-\pi)}$$

$$= \frac{1}{\pi}$$

Now we can calculate $P(C = C_1|TT)$ and $P(C = C_2|TT)$ by using the Bayes' rule. Note that since a coin flip is a Bernoulli trial and there are two types of coins, the following notations hold.

$$P(TT|C) = P(T|C)^2$$
$$P(T|C) = 1 - P(H|C)$$
$$P(C = C_1) = 1 - P(C = C_2) = 0.5$$

Then we can now calculated the likelihood in the following way.

$$P\left(C = C_1 \mid TT\right) = \frac{P\left(C = C_1\right)P\left(TT \mid C = C_1\right)}{P\left(C = C_1\right)P\left(TT \mid C = C_1\right) + P\left(C = C_2\right)P\left(TT \mid C = C_2\right)}$$

$$= \frac{0.5(0.4)^2}{0.5(0.4)^2 + 0.5(0.6)^2} = \frac{16}{52}$$

$$= 1 - P(C = C_2|TT)$$

The posterior expectation of $N$ given $TT$ follows,

$$E(N|TT) = E[E(N|TT, C)|TT]$$
$$= P\left(C = C_1 \mid TT\right)E\left(N \mid C = C_1, TT\right) + P\left(C = C_2 \mid TT\right)E\left(N \mid C = C_2, TT\right)$$
$$= \frac{16}{52}E(N|C_1) + \frac{36}{52}E(N|C_2)$$
$$= \frac{16}{52} * 1/0.6 + \frac{36}{52} * 1/0.4$$
$$= 2.24359$$

**3.**

**(a)** The normal approximation follows,

```
(y_mean <- 1000 * 1/6)
```
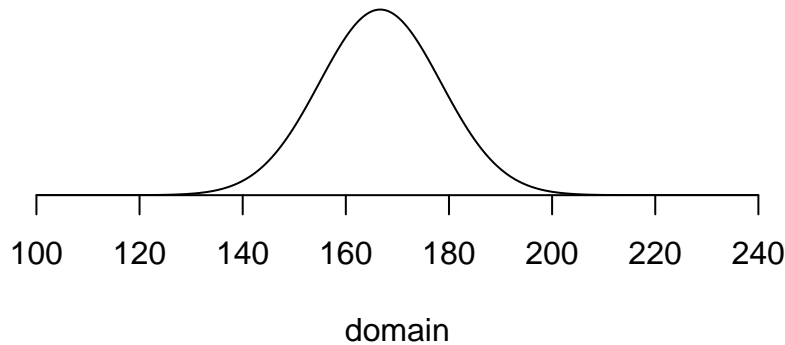
```
## [1] 166.6667
```

```
(y_var <- 1000 * 1/6 * 5/6)
```

```
## [1] 138.8889
```

```
domain <- seq(100, 300, 0.1)
range <- dnorm(domain, y_mean, sqrt(y_var))
plot(domain, range,
    type='l',
    main = 'Distribution of y\nNormal Approximation',
    ylim = c(0, max(range)*1.2), xlim = c(100,240),
    ylab = '', yaxt='n', bty='n', xaxs='i', yaxs='i')
```

## Distribution of y
## Normal Approximation



domain

**(b)**  The 5%, 25%, 50%, 75%, 95% points are calculated likewise.

```r
round(qnorm(0.05, y_mean, sqrt(y_var)))
```

```
## [1] 147
```

```r
round(qnorm(0.25, y_mean, sqrt(y_var)))
```

```
## [1] 159
```

```r
round(qnorm(0.50, y_mean, sqrt(y_var)))
```

```
## [1] 167
```

```r
round(qnorm(0.75, y_mean, sqrt(y_var)))
```

```
## [1] 175
```

```r
round(qnorm(0.95, y_mean, sqrt(y_var)))
```

```
## [1] 186
```

**4.**

**(a)**

$$y|\theta \approx N(1000\theta, 1000\theta(1 - \theta))$$

The approximate prior predictive distribution can be sketched in the following way.

```r
domain <- seq(0, 500, 0.1)

norm_range <- function(domain, theta){
        return(dnorm(domain, 1000*theta, sqrt(1000*theta*(1-theta)))
)}

range_mixed <- 0.25 * norm_range(domain, 1/12) +
        0.5 * norm_range(domain, 1/6) +
        0.25 * norm_range(domain, 1/4)

plot(domain, range_mixed,
    type='l',
    main = 'Prior Predictive Distribution of y\nNormal Approximation',
```

4

```
    ylim = c(0, max(range_mixed)*1.2), xlim = c(30,330),
    ylab = '', yaxt='n', bty='n', xaxs='i', yaxs='i')
```

## Prior Predictive Distribution of y
## Normal Approximation



domain
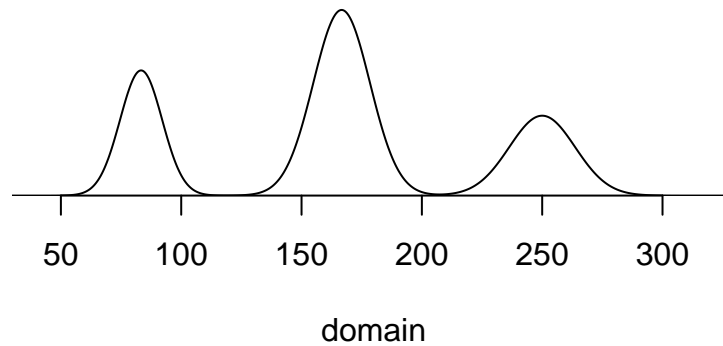
**(b)** Since $p(y)$ has three lumps each taking 25%, 50%, 25% of the whole area under the curve with little overlap, the first 5% point will be the 20% point$(0.25 * 0.2 = 0.05)$.

```
round(qnorm(0.2, 1000*1/12, sqrt(1000*1/12*11/12)))
```

## [1] 76

The 25% point will be between 100 and 150, where the first lump and the second lump meets. We can approximate this by finding the minimum point within the interval.

```
min_point <- min(range_mixed[domain > 100 & domain <150])
round(domain[range_mixed==min_point])
```

## [1] 119

The 50% point will be the midpoint of the second lump. $ 0.25 + 0.5*0.5 = 0.5$.

```
round(qnorm(0.5, 1000*1/6, sqrt(1000*1/6 * 5/6)))
```

## [1] 167

The 75% point will approximately be at the lowest point between the second and the third lump.

```
min_point <- min(range_mixed[domain > 180 & domain <230])
round(domain[range_mixed==min_point])
```

## [1] 207

The 95% point will be located on the 80% point on the third lump. $0.25 + 0.5 + 0.25 * 0.8 = 0.95$.

```
round(qnorm(0.8, 1000*1/4, sqrt(1000*1/4 * 3/4)))
```

## [1] 262

**5.**

**(a)**
$$P(y = k) = \int_0^1 P(y = k \mid \theta) d\theta$$
$$= \int_0^1 \binom{n}{k} \theta^k (1-\theta)^{n-k} d\theta \ \cdots \ Beta(k+1, n-k+1)$$
$$= \binom{n}{k} \frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)}$$
$$= \frac{1}{n+1}$$

**(b)** To show that the posterior mean $\frac{\alpha+y}{\alpha+\beta+n}$ is between the prior mean and the observed relative frequency, we can start with the order between the prior mean and the posterior mean and see if it is also bounded by the observed relative frequency.
First case,

$$Suppose \ \frac{\alpha}{\alpha+\beta} \le \frac{\alpha+y}{\alpha+\beta+n},$$
$$Then, \ \alpha^2 + \alpha\beta + \alpha n \le \alpha^2 + \alpha y + \alpha\beta + \beta y$$
$$\alpha n + ny \le \alpha y + \beta y + ny$$
$$\frac{\alpha+y}{\alpha+\beta+n} \le \frac{y}{n}$$

Second case where $\frac{\alpha}{\alpha+\beta} > \frac{\alpha+y}{\alpha+\beta+n}$ can be shown in the exact same way.
Therefore, the posterior mean is always in between the observed frequency and the prior mean.

**(c)** Uniform distribution is $Beta(1,1)$. Therefore the prior variance is $\frac{1}{12}$. The posterior distribution follows a $Beta(y+1, n-y+1)$ distribution, therefore having the posterior variance $\frac{(y+1)(n-y+1)}{(n+2)^2(n+3)}$.

$$\frac{(y+1)(n-y+1)}{(n+2)^2(n+3)} = \frac{-y^2 + ny + n + 1}{(n+2)^2} \frac{1}{n+3}$$
$$= \frac{-(y - \frac{n}{2})^2 + \frac{n^2}{4} + n + 1}{(n+2)^2} \frac{1}{n+3}$$
$$\le \frac{1}{4 \cdot 3}$$

The first term from the result above is at most $\frac{1}{4}$ and the second term is smaller than $\frac{1}{3}$, therefore the last inequality holds.

**(d)**
$$PriorVar : \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$
$$PostVar : \frac{(\alpha+y)(\beta+n-y)}{(\alpha+\beta+n)^2(n+\alpha+\beta+1)}$$

Since, the posterior variance will on average get smaller as $n \to \infty$ than that of the prior variance, pick minimum $n, y$ such that $n = y = 1$.
Then by algebraically comparing the prior and posterior variance fix $\alpha$(in this case 1), and derive possible values of $\beta$. For example, $n = y = \alpha = 1, \beta = 3$.

**7.**

**(a)**

$$f(y) = \binom{n}{y}$$
$$g(\theta) = (1 - \theta)^n$$
$$u(y) = y$$
$$\phi(\theta) = log(\frac{\theta}{1 - \theta}) \ \cdots \ \text{natural parameter}$$

Then, the uniform prior density of $\phi(\theta), p(\phi) \propto 1 \ for \ \mathbf{R}$ can be transformed to the prior density for $\theta = e^\phi/(1 + e^\phi)$.

$$q(\theta) = p\left(\frac{e^\phi}{1 + e^\phi}\right)\left|\frac{d}{d\theta}\log\left(\frac{\theta}{1 - \theta}\right)\right| \propto \theta^{-1}(1 - \theta)^{-1}$$

**(b)** If $y = 0$ then as $\theta \to 0$, $p(\theta|y) \propto \theta^{-1}(1 - \theta)^{n-1} \to \infty$ Likewise, if $y = n$ then as $\theta \to 1$, $p(\theta|y) \propto \theta^{n-1}(1 - \theta)^{-1} \to \infty$

**8.**

**(a)**

$$y \sim N(\theta, \sigma^2), \ \theta \sim N(\mu_0, \tau_0^2)$$

$$p(\theta) \propto exp(-\frac{1}{2\tau_0^2}(\theta - \mu_0)^2)$$

$$p(\theta|\bar{y}) \propto exp(-\frac{1}{2}(\sum_{i=1}^n \frac{(y_i - \theta)^2}{\sigma^2} + \frac{(\theta - \mu_0)^2}{\tau_0^2}))$$

$$p(\theta|\bar{y}) \propto exp(-\frac{1}{2\tau_1^2}(\theta - \mu_1)^2)$$

where,

$$\mu_1 = \frac{\mu_0/\tau_0^2 + n\bar{y}/\sigma^2}{1/\tau_0^2 + n/\sigma^2} \ \text{and} \ 1/\tau_1^2 = 1/\tau_0^2 + n/\sigma^2$$

$$\mu_0 = 180, \ \tau_0 = 40, \ \sigma = 20, \ \bar{y} = 150$$

$$\theta \mid y \sim N\left(\frac{\frac{180}{40^2} + \frac{150n}{20^2}}{\frac{1}{40^2} + \frac{n}{20^2}}, \frac{1}{\frac{1}{40^2} + \frac{n}{20^2}}\right)$$

**(b)**

$$p(\tilde{y}|y) = \int p(\tilde{y}|\theta)p(\theta|y)d\theta$$

$$\propto \int exp(-\frac{(\tilde{y} - \theta)^2}{2\sigma^2})exp(-\frac{(\theta - \mu_1)^2}{2\tau_1^2})d\theta$$

$$E(\tilde{y}|y) = E(E(\tilde{y}|\theta, y)|y) = E(\theta|y) = \mu_1$$
$$var(\tilde{y}|y) = E(v(\tilde{y}|\theta, y)|y) + var(E(\tilde{\theta}, y)|y)$$
$$= E(\sigma^2|y) + Var(\theta|y)$$
$$= \sigma^2 + \tau_1^2$$

$$\tilde{y} \mid y \sim \text{N} \left( \frac{\frac{180}{40^2} + \frac{150n}{20^2}}{\frac{1}{40^2} + \frac{n}{20^2}}, \frac{1}{\frac{1}{40^2} + \frac{n}{20^2}} + 20^2 \right)$$

**(c)** For the posterior interval,

$$n = 10, \ \mu_0 = 180, \ \tau_0 = 40, \ \sigma = 20, \ \bar{y} = 150$$

$$\theta \mid y \sim \text{N} \left( \frac{\frac{180}{40^2} + \frac{1500}{20^2}}{\frac{1}{40^2} + \frac{10}{20^2}}, \frac{1}{\frac{1}{40^2} + \frac{10}{20^2}} \right)$$

```r
n <- 10
domain <- seq(0,300, 0.1)
tau_1 <- 1/(1/40^2 + n/20^2)
posterior_predictive_var <- tau_1 +20^2
mu_1 <- tau_1*(180/40^2 + 150 * n/20^2)
range <- dnorm(domain,
               mean = mu_1, sd = sqrt(tau_1))
lower_bound <- qnorm(0.025, mu_1, sqrt(tau_1))
upper_bound <- qnorm(0.975, mu_1, sqrt(tau_1))
plot(domain, range,
     type='l',
     main = 'Posterior Distribution of theta\nNormal Approximation',
     ylim = c(0, max(range)*1.2), xlim = c(120,180),
     ylab = '', yaxt='n', bty='n', xaxs='i', yaxs='i')
polygon(c(domain[domain > lower_bound & domain < upper_bound], upper_bound, lower_bound),
        c(range[domain > lower_bound & domain < upper_bound], 0, 0),
        col = 'grey')
```
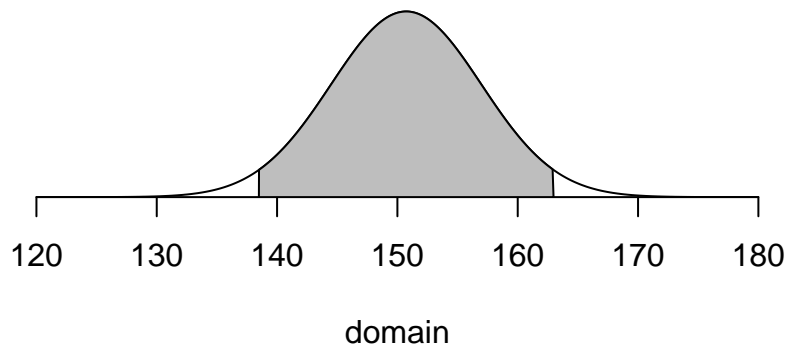
## Posterior Distribution of theta
## Normal Approximation



```r
glue('95% posterior interval for theta: [{round(lower_bound)}, {round(upper_bound)}]')
```

```
## 95% posterior interval for theta: [138, 163]
```

```r
n <- 10
domain <- seq(0,300, 0.1)
tau_1 <- 1/(1/40^2 + n/20^2)
posterior_predictive_var <- tau_1 +20^2
mu_1 <- tau_1*(180/40^2 + 150 * n/20^2)
range <- dnorm(domain,
               mean = mu_1, sd = sqrt(posterior_predictive_var))
```
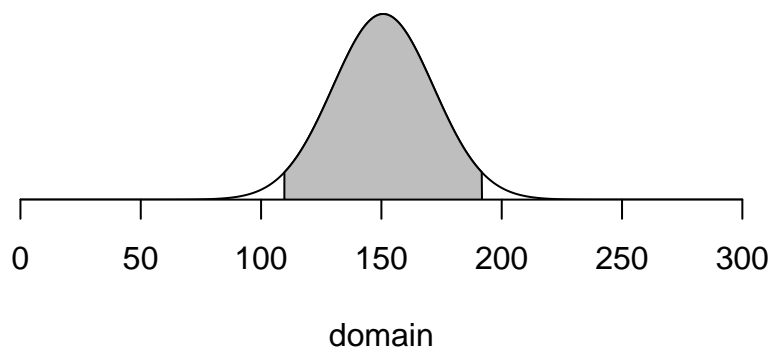
```r
lower_bound <- qnorm(0.025, mu_1, sqrt(posterior_predictive_var))
upper_bound <- qnorm(0.975, mu_1, sqrt(posterior_predictive_var))
plot(domain, range,
     type='l',
     main = 'Posterior Predictive Distribution of tilde y\nNormal Approximation',
     ylim = c(0, max(range)*1.2),
     ylab = '', yaxt='n', bty='n', xaxs='i', yaxs='i')
polygon(c(domain[domain > lower_bound & domain < upper_bound], upper_bound, lower_bound),
        c(range[domain > lower_bound & domain < upper_bound], 0, 0),
        col = 'grey')
```

## Posterior Predictive Distribution of tilde y
## Normal Approximation



domain

```r
glue('95% posterior predictive interval for tilde y: [{round(lower_bound)}, {round(upper_bound)}]')
```

```
## 95% posterior predictive interval for tilde y: [110, 192]
```
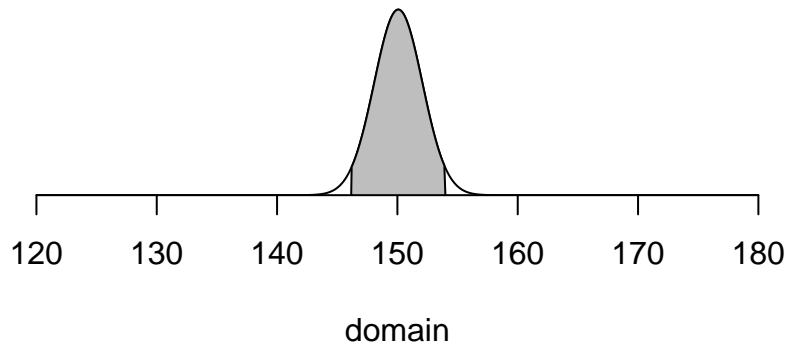
```r
n <- 100
domain <- seq(0,300, 0.1)
tau_1 <- 1/(1/40^2 + n/20^2)
posterior_predictive_var <- tau_1 +20^2
mu_1 <- tau_1*(180/40^2 + 150 * n/20^2)
range <- dnorm(domain,
               mean = mu_1, sd = sqrt(tau_1))
lower_bound <- qnorm(0.025, mu_1, sqrt(tau_1))
upper_bound <- qnorm(0.975, mu_1, sqrt(tau_1))
plot(domain, range,
     type='l',
     main = 'Posterior Distribution of theta\nNormal Approximation',
     ylim = c(0, max(range)*1.2), xlim = c(120,180),
     ylab = '', yaxt='n', bty='n', xaxs='i', yaxs='i')
polygon(c(domain[domain > lower_bound & domain < upper_bound], upper_bound, lower_bound),
        c(range[domain > lower_bound & domain < upper_bound], 0, 0),
        col = 'grey')
```

(d)

## Posterior Distribution of theta
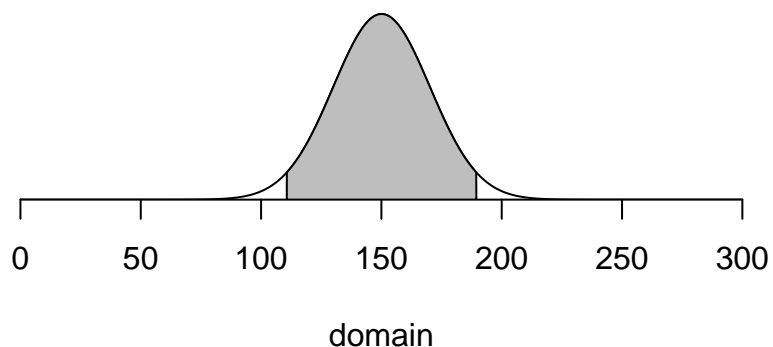## Normal Approximation



domain

```
glue('95% posterior interval for theta: [{round(lower_bound)}, {round(upper_bound)}]')
```

```
## 95% posterior interval for theta: [146, 154]
```

```
n <- 100
domain <- seq(0,300, 0.1)
tau_1 <- 1/(1/40^2 + n/20^2)
posterior_predictive_var <- tau_1 +20^2
mu_1 <- tau_1*(180/40^2 + 150 * n/20^2)
range <- dnorm(domain,
               mean = mu_1, sd = sqrt(posterior_predictive_var))
lower_bound <- qnorm(0.025, mu_1, sqrt(posterior_predictive_var))
upper_bound <- qnorm(0.975, mu_1, sqrt(posterior_predictive_var))
plot(domain, range,
     type='l',
     main = 'Posterior Predictive Distribution of tilde y\nNormal Approximation',
     ylim = c(0, max(range)*1.2),
     ylab = '', yaxt='n', bty='n', xaxs='i', yaxs='i')
polygon(c(domain[domain > lower_bound & domain < upper_bound], upper_bound, lower_bound),
        c(range[domain > lower_bound & domain < upper_bound], 0, 0),
        col = 'grey')
```

## Posterior Predictive Distribution of tilde y
## Normal Approximation



domain

```
glue('95% posterior predictive interval for tilde y: [{round(lower_bound)}, {round(upper_bound)}]')
```

```
## 95% posterior predictive interval for tilde y: [111, 189]
```

**9.**

**(a)**

$$\theta \sim beta(a, b)$$

$$\frac{a}{a+b} = 0.6 \text{ and } \frac{ab}{(a+b)^2(a+b+1)} = 0.3^2$$

Then, we can solve for $a, b$ algebraically.
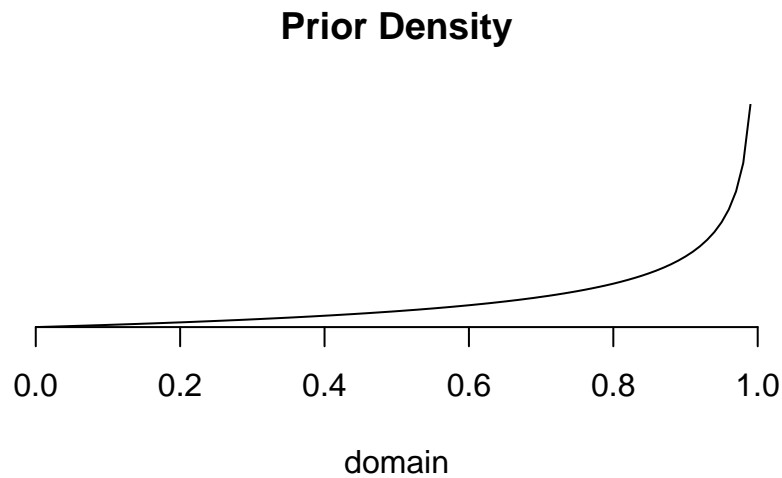
$$a = 1.5b$$

$$\frac{1.5b^2}{2.5^2 b^2 (2.5b + 1)} = 0.09$$

$$b = (1.5/(2.5^2 * 0.09) - 1)/2.5$$

$$= 0.666667 = 2/3$$

$$a = 1$$

```
a = 1
b = 2/3
domain <- seq(0, 1, 0.01)
range <- dbeta(domain, a , b)
plot(domain, range,
     type='l',
     main = 'Prior Density',
     ylab = '', yaxt='n', bty='n', xaxs='i', yaxs='i')
```
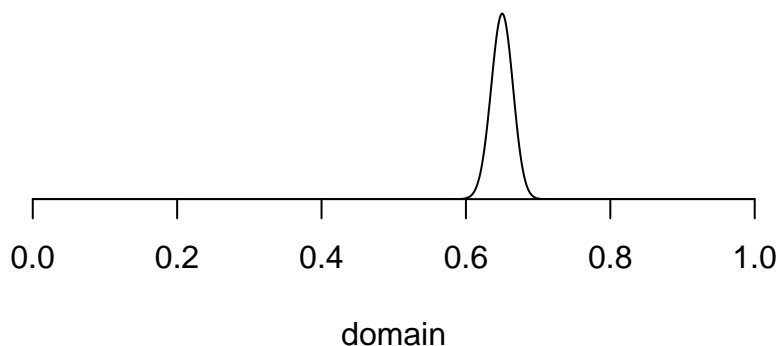
# Prior Density

**(b)**

$$\theta \sim Beta(1, 2/3)$$
$$y|\theta \sim Bin(1000, \theta)$$
$$p(\theta|y) \propto p(\theta)p(y|\theta)$$
$$\propto (1 - \theta^{-1/3}) * \theta^{650}(1 - \theta)^{350}$$
$$= \theta^{650}(1 - \theta)^{350-1/3}$$
$$\theta|y \sim Beta(651, 351 - 1/3)$$
$$E(\theta|y) = 0.6499$$
$$V(\theta|y) = 0.0002$$

```r
a = 651
b = 351-1/3
post_mean <- (a) / (a+b)
post_var <- (a*b) / ((a+b)^2 * (a+b+1))
domain <- seq(0, 1, 0.001)
range <- dbeta(domain, a , b)
plot(domain, range,
     type='l',
     main = 'Posterior Density',
     ylim = c(0, max(range)*1.2), xlim = c(0,1),
     ylab = '', yaxt='n', bty='n', xaxs='i', yaxs='i')
```



```r
glue('Posterior Mean: {post_mean}, Posterior Var: {post_var}')
```

```
## Posterior Mean: 0.649916805324459, Posterior Var: 0.000226919831929496
```

**(c)**   Taking $y = 650, n = 1000$ from **(b)**. Prior will be plotted in a blue line where as the posterior will be in a solid brown.

```r
y <- 650
n <- 1000
domain <- seq(0, 1, 0.001)

example_9_plotter <- function(prior_alpha, prior_beta) {
      prior_a <- prior_alpha
      prior_b <- prior_beta
      prior_range <- dbeta(domain, prior_a, prior_b)
```

```
        post_a <- prior_a + y
        post_b <- n-y + prior_b
        post_range <- dbeta(domain, post_a, post_b)

        plot(domain, post_range,
        type='l',
        main = glue('Prior({prior_a},{prior_b}), Posterior({post_a},{post_b})'),
        ylim = c(0, max(prior_range, post_range)*1.2), xlim = c(0,1),
        ylab = '', yaxt='n', bty='n', xaxs='i', yaxs='i', col='brown')
        lines(domain, prior_range,
        type='l' ,col='blue')
}

par(mar=c(4,3,3,1))
par(mfrow=c(4,2))
example_9_plotter(1,1)
example_9_plotter(10,10)
example_9_plotter(100,10)
example_9_plotter(500,10)

example_9_plotter(10,100)
example_9_plotter(10,500)
example_9_plotter(10,1000)
example_9_plotter(500,500)
```
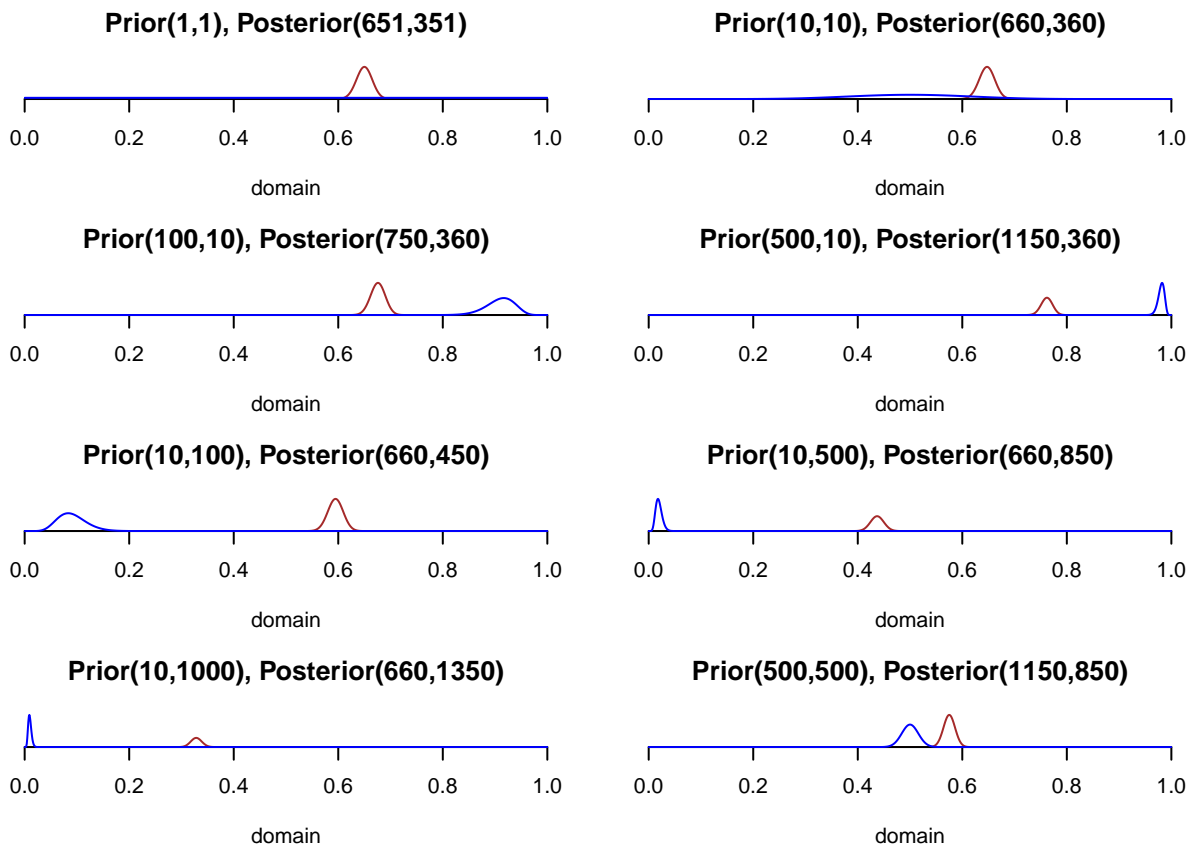
**10.**

**(a)**

$$\text{Likelihood: } p(obs|N) = 1/N \ if \ N \geq 203 \ o.w. \ 0$$
$$\text{prior: } p(N) = (1/100)(99/100)^{N-1} \ for \ N \in \mathbf{N}$$
$$\text{posterior: } p(N|obs) \propto (99/100)^{N} * 1/N \ for \ N \geq 203$$

**(b)** Computer approximation for the posterior mean and variance of **(a)** can be done in the following way. First, we need to calculate the normalizing constant $C$.

| $N|obs$ | mass |
|---------|------|
| 203 | C*(99/100)^203 / 203 |
| 204 | C*(99/100)^204 / 204 |
| 205 | C*(99/100)^205 / 205 |
| ... | ... |
| k | C*(99/100)^k / k |

```
domain <- seq(203, 10000, 1)
improper_post_prob <- function(x) {
        return((99/100)^x/x)
}

norm_constant <- 1/sum(improper_post_prob(domain))
glue('Normalizing Constant C is: {norm_constant}')
```

```
## Normalizing Constant C is: 21.4682901123524
```

```
post_prob <- function(x) {
        return(norm_constant * (99/100)^x /x)
}
post_matrix <- cbind(domain, post_prob(domain))
post_mean <- sum(post_matrix[,1] * post_matrix[,2])
post_var <- sum(post_matrix[,1]^2 * post_matrix[,2]) - post_mean^2
glue('Posterior mean and standard deviation: ({round(post_mean)}, {sqrt(post_var)})')
```

```
## Posterior mean and standard deviation: (279, 79.9645754036781)
```

**(c)** If we take the prior distribution as $p(N) \propto 1/N$ for $N \geq 203$, the posterior distribution is proportional to $1/N^2$. The infinite sum of $1/N^2$ is also known as the Basel problem in mathematical analysis. Thanks to Euler, we can use the following fact. $\sum_{n=1}^{\infty} 1/n^2 = \pi^2/6$.

$$\sum_{n=1}^{\infty} 1/n^2 = \pi^2/6 = \sum_{n=1}^{202} 1/n^2 + \sum_{n=203}^{\infty} 1/n^2$$

```
domain <- 1:202
(const <- sum(1/domain^2))
```

```
## [1] 1.639996
```

The value of the normalizing constant follows.

$$C = \sum_{n=203}^{\infty} 1/n^2 = \pi^2/6 - 1.639996 = 0.0049$$

```
(pi^2/6 -const)
```

## [1] 0.004938262

If the posterior distribution is divided by $C$ then we have a proper probability distribution.

$$p(N|obs) = \frac{1/N^2}{0.0049} \text{ where } N \geq 203$$

The posterior mean and the standard deviation for N follows.

```
domain <- 203:10000
post_prob <- function(x) {
        return( (1/x^2)/0.0049)
}
post_mean <- sum(domain * post_prob(domain))
post_var <- sum(domain^2 * post_prob(domain)) - post_mean^2
glue('Posterior mean and standard deviation: ({round(post_mean)}, {sqrt(post_var)})')
```

## Posterior mean and standard deviation: (796, 1168.8539940591)

Since the posterior variance is ridiculously large, we know that in this particular case(guessing the total number of things based on one observation) presuming a non-informative prior is not a good practice.

**11.**

**(a)**

$$\mathbf{y} : \text{vector of observed data}$$

$$p(\mathbf{y}|\theta) \propto \prod_{\forall i} 1/(1 + (y_i - \theta)^2) \ \cdots \ y_i \in \mathbf{y}$$

$$p(\theta|\mathbf{y}) = \frac{0.01 * \prod_{\forall i} 1/(1 + (y_i - \theta)^2)}{C}$$

where $C$ is the normalizing constant.

We can calculate the area under the `unnormalized_posterior_density` over the `domain` by using the `integrate.xy` function. And that value will be $C$.

```
domain <- seq(0,100,0.01)
obs <- c(43,44,45,46.5,47.5)
caucy_proportional_likelihood <- function(observations, theta) {
        res <- 1
        for (observation in observations){
                iter_res <- dcauchy(theta, observation,1)
                res <- res * iter_res
        }
        return(res)
}
unnormalized_posterior_density <- 0.01 * caucy_proportional_likelihood(obs, domain)
normalizing_constant <- integrate.xy(domain, unnormalized_posterior_density)
glue('Normalizing Constant C: {normalizing_constant}')
```

## Normalizing Constant C: 3.41835903112125e-07

```
normalized_posterior_density <- unnormalized_posterior_density / normalizing_constant
auc_normalized_posterior_density <- integrate.xy(domain, normalized_posterior_density)
glue('Area Under Normalized posterior density: {auc_normalized_posterior_density}')
```

```
## Area Under Normalized posterior density: 1
```

```r
plot(domain, normalized_posterior_density,
     type='l',
     main = 'Normalized Posterior Density of theta',
     ylim = c(0, max(normalized_posterior_density)*1.2), xlim = c(40, 50),
     ylab = '', yaxt='n', bty='n', xaxs='i', yaxs='i')
```
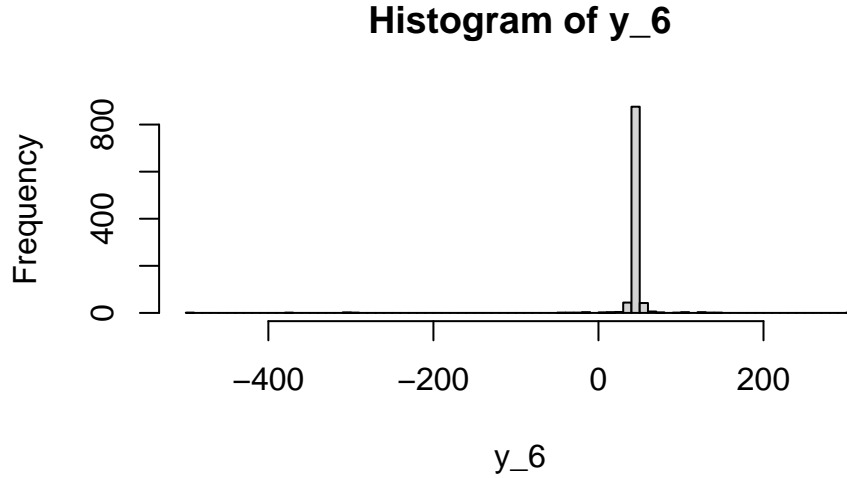
**Normalized Posterior Density of theta**
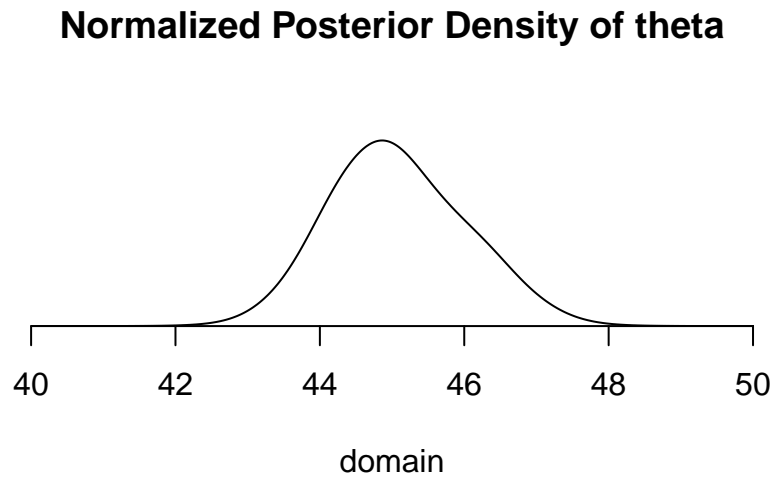


```r
set.seed(111)
theta_samples <- sample(domain, 1000, prob = normalized_posterior_density, replace = TRUE)
hist(theta_samples, breaks = 100)
```

**(b)**

**Histogram of theta_samples**
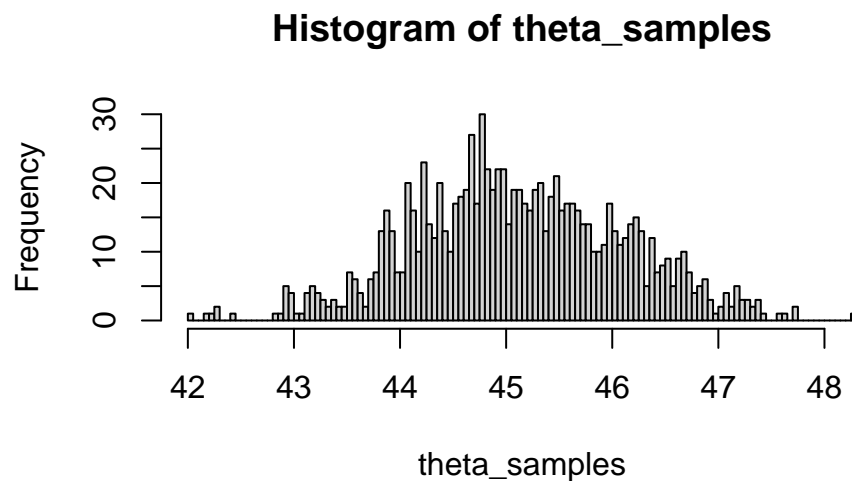


```r
y_6 <- rcauchy(1000, theta_samples,1)
hist(y_6, breaks = 100)
```

**(c)**

## Histogram of y_6



**12.**

Jeffreys' prior($p(\theta)$) of $\theta$ is proportional to the square root of the Fischer information($I(\theta)$) in a single parameter case. Since the likelihood function($p(y|\theta) = \theta^y exp(-\theta)/y!$) is twice differentiable, we can easily find the Jeffrey's prior in the following way.

$$p(\theta) \propto \sqrt{I(\theta)}$$
$$= \sqrt{-E[\frac{d^2}{d\theta^2}p(y|\theta)|\theta]}$$
$$= \sqrt{-E[-\frac{y}{\theta^2}|\theta]}$$
$$= \sqrt{1/\theta}$$
$$= \theta^{-1/2}$$

This is the same pdf for the $Gamma(1/2, 0)$ distribution.

**13.**

**(a)** Let the prior distribution follow a Gamma distribution($Gamma(a, b)$) and the likelihood function follow a Poisson distribution($Pois(\theta)$).

$$f(\theta) = \frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta}$$

$$f(\mathbf{x}|\theta) = \prod_{i=1}^{n}\frac{\theta^{x_i}e^{-\theta}}{x_i!} \propto \theta^{n\bar{x}}exp(-n\theta)$$

$$f(\theta|\mathbf{x}) \propto \theta^{n\bar{x}+a-1}exp(-(b+n)\theta)$$

Therefore, the posterior distribution follows a Gamma distribution($Gamma(n\bar{x}+a, b+n)$). If we use a neutral noninformative prior Gamma distribution we can set $(a, b) = (1/3, 0)$ from the study of Jouni Kerman's *Neutral noninformative and informative conjugate beta and gamma prior distributions*. The posterior distribution will be $Gamma(1/3 + n\bar{x}, n)$.

```
domain <- seq(0, 100, 0.01)
accidents <- c(24,25,31,31,22,21,26,20,16,22)
n <- length(accidents)
```

17

```
post_density <- dgamma(domain, n*mean(accidents), n)
theta_samples <- sample(domain, 1000, prob = post_density)
y_1986 <- rpois(1000, theta_samples)
lower_bound <- sort(y_1986)[25]
upper_bound <- sort(y_1986)[975]
glue('95% predictive interval for fatal accidents in 1986: [{lower_bound}, {upper_bound}]')
```

```
## 95% predictive interval for fatal accidents in 1986: [13, 36]
```

**(b)**

$$\text{flown miles} = \text{passenger deaths}/\text{death rate} * 10^8$$

```
passenger_deaths <- c(734,516,754,877,814,362,764,809,223,1066)
death_rate <- c(0.19,0.12,0.15,0.16,0.14,0.06,0.13,0.13,0.03,0.15)
flown_miles <- passenger_deaths / death_rate * 10^8
```

$$If, \ k_i : \text{flown miles in i-th year}, \ \theta : \text{accident rate per mile}$$

$$Then, \ x_i|k_i, \theta \sim Poisson(k_i\theta)$$

If we plug in $k_i\theta$ to $\theta$ in the likelihood function in **(a)**, our posterior density will be,

$$f(\theta|obs) \propto \theta^{n\bar{x}+a-1} exp(-(b + n\bar{k})\theta)$$

where the parameters being $Gamma(a + n\bar{x}, b + n\bar{k})$. When we again use the neutral noninformative Gamma as a prior, we end up with the posterior distribution,

$$obs|\theta \sim Gamma(1/3 + n\bar{x}, n\bar{k}).$$

The posterior predictive distribution is $\tilde{x}|\theta \sim Poisson(\tilde{k}\theta)$ where $\tilde{k} = 8 * 10^{11}$

```
k_bar <- n*mean(flown_miles)
theta_samples <- rgamma(1000, 1/3+ n*mean(accidents))/ k_bar
y_1986 <- rpois(1000, theta_samples * 8*10^(11))
lower_bound <- sort(y_1986)[25]
upper_bound <- sort(y_1986)[975]
glue('95% predictive interval for fatal accidents in 1986: [{lower_bound}, {upper_bound}]')
```

```
## 95% predictive interval for fatal accidents in 1986: [22, 47]
```

```
theta_samples <- rgamma(1000, 1/3+ n*mean(passenger_deaths))/ n
y_1986 <- rpois(1000, theta_samples)
lower_bound <- sort(y_1986)[25]
upper_bound <- sort(y_1986)[975]
glue('95% predictive interval for Passenger deaths in 1986: [{lower_bound}, {upper_bound}]')
```

**(c)**

```
## 95% predictive interval for Passenger deaths in 1986: [635, 744]
```

```
k_bar <- n*mean(flown_miles)
theta_samples <- rgamma(1000, 1/3+ n*mean(passenger_deaths))/ k_bar
y_1986 <- rpois(1000, theta_samples*8*10^11)
```

```
lower_bound <- sort(y_1986)[25]
upper_bound <- sort(y_1986)[975]
glue('95% predictive interval for Passenger deaths in 1986: [{lower_bound}, {upper_bound}]')
```

**(d)**

```
## 95% predictive interval for Passenger deaths in 1986: [906, 1036]
```

**(e)** The model in **(b)** seems to be most appropriate in terms of common sense. If we use an analogy to a situation where we need to buy used cars, we would expect a vehicle with a high mileage to have a greater number of damage compared to a brand new car. This logic can be roughly applied to the subject matter. Also, models in **(c)**,**(d)** does not quite satisfy the Poisson case, since passenger deaths are caused in large bulks depending on the number of accidents.

**16.**

**(a)**

$$p(y) = \int p(y \mid \theta)p(\theta)d\theta$$

$$= \int_0^1 \binom{n}{y} \theta^y(1-\theta)^{n-y}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta$$

$$= \frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}d\theta$$

$$= \frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(y+\alpha)\Gamma(n-y+\beta)}{\Gamma(n+\alpha+\beta)}$$

#### (b) We need to see what condition satisties $p(y) = C$ where $C$ is a constant. Since the gamma functions with the term $y$ is what makes $p(y)$ vary, we can examine only those terms and find a condition to cancel them out.

$$\frac{\Gamma(y+\alpha)\Gamma(n-y+\beta)}{\Gamma(y+1)\Gamma(n-y+1)}$$

If $\alpha = \beta = 1$ then $p(y)$ becomes a constant.

**17.**

**(a)** Set $u = \sigma^2$. Then by transformation of random variables we have, $p(\sigma^2) = p(u) = p(\sqrt{u})d\sqrt{u}/du = p(\sigma) * 1/2 * u^{-1/2} \propto p(\sigma)/\sigma \propto 1/\sigma^2$ since $p(\sigma) \propto 1/\sigma$.

**(b)** Suppose the 95% hpd for $p(\sigma|obs)$ is $(\sqrt{a}, \sqrt{b})$ which implies $a < b$. We know that $p(\sigma|obs) \propto \sigma^{-1-n}\exp(-nv/2\sigma^2)$. And if we plug in the supposed interval we get,

$$a^{-1/2-n/2}\exp(-nv/2a) = b^{-1/2-n/2}\exp(-nv/2b)$$

$$(-1/2 - n/2)\log a - nv/2a = (-1/2 - n/2)\log b - nv/2b.$$

Suppose the 95% hpd for $p(\sigma^2|obs)$ is $(a, b)$. Then we have,

$$a^{-1-n/2}\exp(-nv/2a) = b^{-1-n/2}\exp(-nv/2b)$$

$$(-1 - n/2)\log a - nv/2a = (-1 - n/2)\log b - nv/2b.$$

Resulting in,

$$-1/2\log a = -1/2\log b \iff a = b$$

$$\text{This contradicts } a < b.$$

Therefore, the 95% highest posterior density region for $\sigma^2$ is not the same as the region obtained by squaring the endpoints of a posterior interval for $\sigma$.

**20.**

**(a)**

$$p(\theta|y \geq 100) \propto p(y \geq 100|\theta)p(\theta)$$
$$\propto (1 - F(100|\theta))p(\theta)$$
$$\propto \exp(-100\theta)\theta^{\alpha-1}\exp(-\beta\theta)$$
$$p(\theta|y \geq 100) = \text{Gamma}(\theta|\alpha, \beta + 100)$$
$$\text{Posterior (mean, variance)} = (\frac{\alpha}{\beta + 100}, \frac{\alpha}{(\beta + 100)^2})$$

**(b)**

$$p(\theta \mid y = 100) \propto p(y = 100 \mid \theta)p(\theta)$$
$$\propto \theta \exp(-100\theta)\theta^{\alpha-1}\exp(-\beta\theta)$$
$$p(\theta \mid y = 100) = \text{Gamma}(\theta \mid \alpha + 1, \beta + 100)$$
$$\text{Posterior (mean, variance)} = (\frac{\alpha + 1}{\beta + 100}, \frac{\alpha + 1}{(\beta + 100)^2})$$

**(c)** **(a)** is about having a prior knowledge on the possible range of values of $y$ for the likelihood function without observing an exact value. **(b)** is actually observing a single data point of $y$. Intuitively the first case, seems to contain a slightly more valuable information compared to the second case in terms of getting to know the possible and impossible values of $y$. From the second case, even though we observe a realization of $y$ we lack information of impossible values compared to the first case.

However this does not contradict (2.8) since it states that the variance of $\theta|y$ decreases *on average* given more information. In other words, $E(\text{var}(\theta \mid y) \mid y \geq 100) \neq E(\text{var}(\theta \mid y) \mid y = 100)$.

## [2] Probability of a girl birth given placenta previa

We can set up the variable in the following way.

$$y : \text{event of female birth; 437}$$
$$n : \text{total number of births; 980}$$
$$\theta : \text{probability of placenta previa}$$

Then we can construct a Binomial model to see whether the claim that the proportion of female births in the population of placenta previa is less than 0.485.

**Prior : Uniform(0,1)**

First, we can use a Uniform(0,1) non-informative prior for $\theta$. Since, Uniform(0,1) distribution is a special case of the Beta distribution($Beta(\alpha, \beta)$), we can directly derive the posterior distribution by conjugacy.

$$\theta|y \sim Beta(437 + 1, 980 - 437 + 1)$$
$$\sim Beta(438, 544)$$

The mean and variance of the posterior distribution can be derived mathematically.

$$E[\theta|y] = \frac{438}{438 + 544} \approx 0.4460$$

$$Var(\theta|y) = \frac{438 * 544}{(438 + 544)^2(438 + 544 + 1)} \approx 0.0002513602 \approx 0.0159^2$$

The 95% credible interval can be computed in the following way.

```
(lower_bound <- qbeta(0.025, 438, 544))
```

## [1] 0.4150655

```
(upper_bound <- qbeta(1-0.025, 438, 544))
```

## [1] 0.4771998

```
domain <- seq(0,1,0.001)
image <- dbeta(domain, 438, 544)
plot(domain, image,
     main='Posterior distribution\nBeta(438,544)',
     frame.plot = FALSE,
     type = 'l',
     xlim = c(0.3, 0.6), ylim = c(0, max(image)*1.1),
     ylab = '', yaxt='n', bty='n', xaxs='i', yaxs='i'
     )
polygon(c(domain[domain > lower_bound & domain < upper_bound], upper_bound, lower_bound),
        c(image[domain > lower_bound & domain < upper_bound], 0, 0),
        col = 'grey')
abline(v = 0.485, lty=2, col='red')
text(0.5,6, labels = '0.485', col = 'red', cex = 0.7)
```



**Posterior distribution**
**Beta(438,544)**
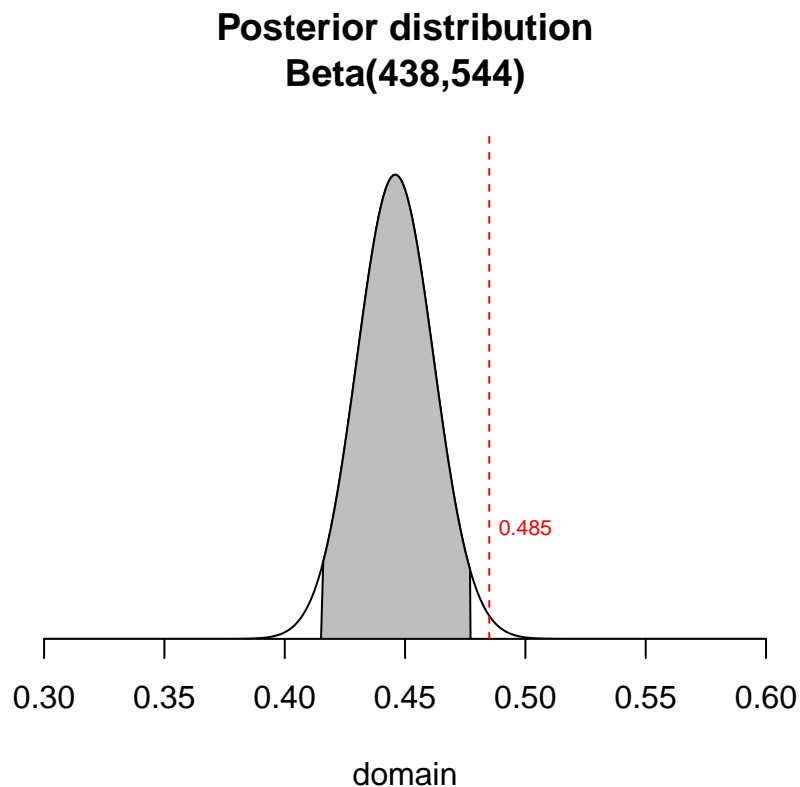
**Prior: Beta(Non-uniform)**

We can set up a function for exploration of various Beta priors. Increasing $a$ will send the mean of the posterior rightward, and increasing $b$ will leftward.

```
set.seed(111)
post_interval <- function(domain, a, b) {
        post_mean <- (437+a)/(980 + a+b)
```

```
        theta_samples <- rbeta(1000, 437+a, 543+b)
        lb <- sort(theta_samples)[25]
        ub <- sort(theta_samples)[975]
        glue('=================  Prior beta({a},{b})  =================\n
        Posterior Mean: {round(post_mean,4)}\r
        95% Posterior Interval: [{round(lb,4)},{round(ub,4)}]\n\n')
}

post_interval(domain, 3,3)
```

```
## =================  Prior beta(3,3)  =================
##
##        Posterior Mean: 0.4462
##        95% Posterior Interval: [0.4154,0.4771]
```
```
post_interval(domain, 200,30)
```

```
## =================  Prior beta(200,30)  =================
##
##        Posterior Mean: 0.5264
##        95% Posterior Interval: [0.4991,0.555]
```
```
post_interval(domain, 20,300)
```

```
## =================  Prior beta(20,300)  =================
##
##        Posterior Mean: 0.3515
##        95% Posterior Interval: [0.3238,0.3772]
```
```
post_interval(domain, 300,400)
```

```
## =================  Prior beta(300,400)  =================
##
##        Posterior Mean: 0.4387
##        95% Posterior Interval: [0.4149,0.4621]
```

**Prior: Nonconjugate prior**

We can think of a unimodal triangle with the tip at 0.485.

```
pos_slope_domain <- domain[domain < 0.485]
neg_slope_domain <- domain[domain >= 0.485]

pos_slope_density <- pos_slope_domain * 2 / (0.97*0.485)
neg_slope_density <- (0.97-neg_slope_domain) * 2 / (0.97*0.485)
image <- c(pos_slope_density, neg_slope_density)



post_density <- image * choose(980, 437) * domain^437 * (1-domain)^543
norm_const <- integrate.xy(domain, post_density)

proper_post_density <- (post_density/norm_const)


plot(domain, proper_post_density,
    main='Posterior distribution(brown)\nPrior distribution(blue)',
```
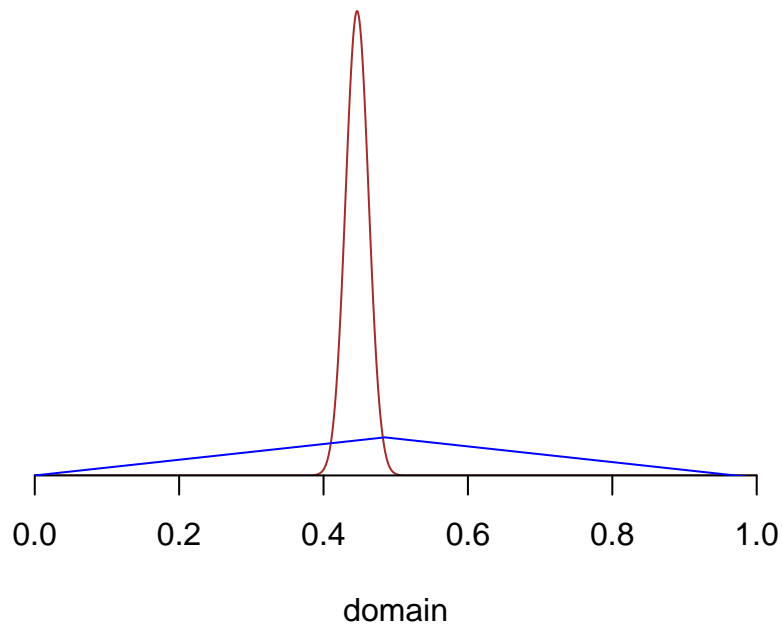
```
    frame.plot = FALSE,
    type = 'l',
    xlim = c(0,1), ylim = c(0, max(image, proper_post_density)*1.1),
    ylab = '', yaxt='n', bty='n', xaxs='i', yaxs='i', col='brown'
    )

lines(domain, image, col = 'blue')
```
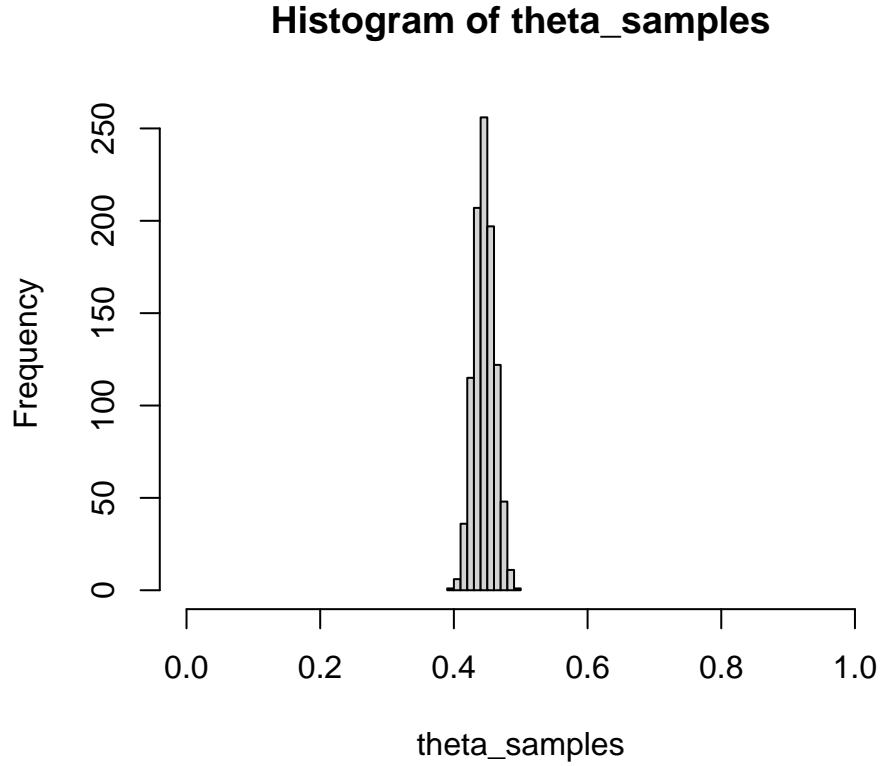
## Posterior distribution(brown)
## Prior distribution(blue)



domain

```
theta_samples <- sample(domain, 1000, prob=proper_post_density, replace = TRUE)
hist(theta_samples, breaks=10, xlim = c(0,1))
```

## Histogram of theta_samples



As we can see in this case, the data dominates the prior distribution.

Since, the upper bound is less than 0.485, we can conclude that the probability of a girl birth given placenta previa is less than 0.485.

## [3] Effect of prior distribution in a Binomial model

For a beta prior and a binomial likelihood, the posterior distribution also follows a beta distribution.

$$\theta \sim Beta(a, b)$$
$$y|\theta \sim Binomial(n, \theta)$$
$$\theta|y \sim Beta(a + y, b + n - y)$$

In a Binomial model like the above, the Beta prior is valuable since it has conjugacy and possibility of intuitive interpretation for parameters $a, b$. Conjugacy is a property of a prior distribution making the posterior distribution belong in the same distribution family and this is shown through out **Section [1]** of this homework. The intuitive interpretation of the prior parameters is possible in the way in which $a$ representing the number of success and $b$ the number of failure.

## [4] Derivation of Jefferey's prior for Binomial distribution

$$Jeffreys' \ prior : p(\theta) \propto \sqrt{det(I(\theta))}$$
$$y|\theta \sim Bin(n, \theta)$$
$$log(f(y|\theta)) = l(\theta) = log(\binom{n}{y}) + ylog\theta + (n - y)\log(1 - \theta)$$

24

$$I(\theta) = -E[\frac{d^2 l(\theta)}{d\theta^2}]$$
$$= E[y/\theta^2 + (n-y)/(1-\theta)^2]$$
$$= E[E[y/\theta^2 + (n-y)/(1-\theta)^2|\theta]]$$
$$= E[n\theta/\theta^2 + (n-n\theta)/(1-\theta)^2]$$
$$= n/\theta + n/(1-\theta)$$
$$= n/\theta(1-\theta)$$
$$p(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2}$$
$$\theta \sim Beta(1/2, 1/2)$$